DOING WITHOUT REPRESENTING?

Andy Clark and Josefa Toribio

SYNTHSESE 101, 1994: 401-431.

SPECIAL ISSUE ON

CONNECTIONISM AND THE FRONTIERS OF ARTIFICIAL INTELLIGENCE

Abstract

Connectionism and classicism, it generally appears, have at least this much in common: both place some notion of internal representation at the heart of a scientific study of mind. In recent years, however, a much more radical view has gained increasing popularity. This view calls into question the commitment to internal representation itself. More strikingly still, this new wave of anti-representationalism is rooted not in 'armchair' theorizing but in practical attempts to model and understand intelligent, adaptive behavior. In this paper we first present, and then critically assess, a variety of recent anti-representationalist treatments. We suggest that so far, at least, the sceptical rhetoric outpaces both evidence and argument. Some probable causes of this premature scepticism are isolated. Nonetheless, the anti-representationalist challenge is shown to be both important and progressive insofar as it forces us to see beyond the bare representational / non-representational dichotomy and to recognize instead a rich continuum of degrees and types of representationality.

0. From Heidegger to Artificial Insects.

Cognitive Science, it has often seemed, is agreed on at least this: that at the heart of a scientific understanding of cognition lies one crucial construct, the notion of internal representation. Different approaches have proposed to unpack this notion in different ways. So-called classicists (e.g. Newell & Simon, 1972, Fodor & Pylyshyn, 1988) favouring a vision in which internal representations exhibit a quasi-linguistic, combinatorial structure. Connectionists (e.g. Smolensky, 1988) favouring instead a mode of representation (vector coding in a high dimensional space) which arguably differs in respect of the basic processing operations, and the kind of semantic structure captured by the computational vehicles of representation.

Both camps, however, retained the fundamental idea of inner computational states acting as the vehicles of specific contents -that is to say, they retained the very idea of internal representation, but held differing views concerning its form. In recent years, however, a much more radical view has gained ground. This view calls into question the commitment to internal representation itself. Cognition, according to this view, need not involve the creation and manipulation of anything deserving the name of 'internal representations' at all. Behind such a radical view lies (often) a more general distrust of the way the notion of internal representation is used to create 'mind' as a special and seperate arena largely insulated from issues concerning embodiment and environmental surround.

Such a distrust has its clearest roots in the Heideggerian rejection of Cartesianism. This rejection involved an attempted inversion of Descartes' COGITO. Instead of invoking as the starting point of philosophical analysis a subject representing the world, Heidegger emphasizes the everyday skills and practices that constitute our being-in-the-world as the essence of cognition (Cfr. Heidegger, 1962, p. 46 and 254). This conceptually based rejection of Representationalism is, however, not our main focus in what follows. Instead, our interest lies with the growing distrust of Representationalism displayed by active practicioners of Cognitive Science; a distrust rooted in the practical attempt to model and understand intelligent, adaptive behavior.

Our claim will be that the empirically driven anti-representationalist vastly overstates her case. Such overstatement is rooted, we suggest, in an unwarranted conflation of the fully general notion of representation with the vastly more restrictive notions of explicit representation and/or of representations bearing intuitive, familiar contents. This is compounded by a serious problem of illustration. The empirically

driven anti-representationalist invokes superficially compelling case studies of complex but representation-free behavior. But these case studies, on closer examination, are compromised by a failure to address the right type of problem viz, the range of cases in which ambient environmental information is (*prima facie*) insufficient to guide behavior.

In what follows we present a variety of arguments and illustrations drawn from the radical anti-representationalist camp and show how these arguments trade on the twin infelicities just described. Connectionist approaches, we suggest, fall into an important category which the anti-representationalist case fails adequately to address insofar as they are deeply representational, yet eschew the use of traditional explicit representations.

The strategy is as follows. Section 1 addresses the issue of what makes an inner state count as a representation. Section 2 rehearses one form of anti-representationalist challenge: Rodney Brooks work on mobile robots. Section 3 goes on to address Randall Beers impressive work on leg controllers for robot insects. We are then in a position (Section 4) to formulate and assess specific recent anti-representationalist arguments. Such arguments, we contend, either confound different notions of representation or involve unwarranted projections from simple to more 'representation-hungry' cases. We end (section 5) by reviewing the dialectic and suggesting that it may be fruitful to stop thinking in terms of a dichotomy (representation / no-representation) and instead imagine a continuum of cases.

1. Representation: Chunky or Smooth?

Connectionists and classicists (e.g. Smolensky, 1988, Fodor and Pylyshyn, 1988) had at least this much in common: both camps held firm to the foundational role of the notion of internal representation in explaining intelligent behavior. The debate between the classicist and the connectionist concerned not the existence or explanatory role of representations, so much as their form and properties. Classicists opted for a 'quasi-sentential' (P. M. Churchland, 1989) approach in which key contents were tokenable as strings of symbols, and were operated upon by a 'read / write / copy' architecture. By contrast, connectionists opted for an architecture in which representation and processing were deeply intertwined, and strings of symbols participating in 'cut and paste' processing were replaced by episodes of vector to vector transformation in high dimensional state spaces (for extended explanation and description see e.g., Rumelhart,

McClelland and the PDP Research Group, 1986, vol I & II, Clark, 1989, 1993, P. M. Churchland, 1989).

The key disagreement could be (and often was) expressed in terms of explicit vs. implicit representation of information (see e.g., Cleeremans, 1993). The classicist vision of explicitness as involving the tokening of strings of symbols able to participate in a real cut and paste ('literally compositional') computational economy was eschewed by the connectionist whose mode of representation typically involved much less transparent and easily manipulable elements (see van Gelder, 1990, 1991). In a fairly intuitive sense, it seemed correct to say of a trained-up network that it could embody knowledge about a domain without explicitly representing the knowledge (at least as any kind of syntactically structured item in a declarative code). Instead, networks embody a powerful kind of 'knowing how' -a knowing how which is now extended into domains once depicted as requiring explicit syntactically structured representation of knowledge (see e.g., the debate over knowledge of the past tense (Rumelhart & McClelland, 1986, Pinker & Prince, 1988). Thus, if by 'explicit representation' we mean something like 'representation as a symbol string in a compositional declarative code', then we may indeed cast the connectionist / classicist debate as a debate concerning the need to invoke explicit representations in explanations of cognitive phenomena.

Explicit, syntactically structured representations and connectionist, distributed representations are thus both species of a more general notion of internal representation. If we do not wish to simply identify representation with explicit symbolization in a classical cognitive architecture, it behoves us to offer at least a sketch of this more general notion. Fortunately, this is not hard to do. We can borrow quite a neat account from e.g. Haugeland (Haugeland, 1991). Haugeland depicts a system as representational just in case:

> (1)It must co-ordinate its behaviors with environmental features which are not always 'reliably present to the system' via some signal.
>
> (2)It copes with such cases by having something else (other than the signal directly received from the environment) 'stand in' and guide behavior in its stead.
>
> (3)That 'something else' is part of a general representational scheme which allows the 'standing in' to occur systematically and allows for a variety of related representational states.
>
> (see Haugeland, 1991, p. 62)

Point (1) rules out cases where there is no 'stand in' at all, and in which the environmental feature (via a 'detectable signal') itself controls the behavior. Thus "plants that track the sun with their leaves need not represent it or its position because the tracking can be guided directly by the sun itself" (Op. cit., p. 62). Point (2) identifies as a representation anything which 'stands in' for the relevant environmental feature. But point (3) narrows the class to include only stand-ins which figure in a larger scheme of standing-in, thus ruling out e.g., gastric juices as full-blooded representations of future food (Op. cit, p. 62).

A Haugeland-style account thus depicts connectionist (implicit) representation and classical (explicit) representation as involving different kinds of overall representational scheme, but nonetheless, as both trading on the basic notion of internal representation. It will be our contention, in what follows, that recent attempts to construct general anti-representationalist arguments fail. And that they fail primarily because they apply only to a specific sub-class of types of internal representation viz those which posit explicit, 'moveable' tokens as the bearers of semantic content. Stripped of the excess baggage of the 'Language of Thought' hypothesis (Fodor, 1975, 1987), the idea of internal representation remains an essential tool for understanding the behavior of intelligent systems.

2. Mobots and Activity-Based Decomposition.

Among the most influential of the recent wave of anti-representationalists are Rodney Brooks and the so-called 'moboticists'. A 'mobot' is a mobile robot capable of functioning in a messy, unpredictable real environment (such as an office). In setting out to design and build such 'creatures', Brooks and others have deliberately turned around much of the traditional methodology of Artificial Intelligence research. Instead of focusing on isolated aspects of the sophisticated cognitive competence of human agents (aspects like chess-playing, language-understanding, past-tense production, etc), Brooks believes we should focus on much simpler complete systems: systems aspiring to roughly the level of insect intelligence. One reason for this approach is a deep scepticism concerning our current abilities to find the right decomposition of human-level intelligence into sub-tasks / modules, etc. Relatedly, the issue of the correct interfaces between such modules is seen as equally opaque to our current understanding. The initial stress on simpler whole intelligences is meant to provide a tractable domain in which to begin to understand biologically realistic kinds of

decomposition and interface. But once we begin to do so, Brooks claims, we are rapidly led to an unexpected conclusion. It is that

> "when we examine very simple level intelligence we find that explicit representations and models of the world simply get in the way. It turns out to be better to use the world as its own model"
>
> Brooks, 1991, p. 140

From this, Brooks moves to a radical hypothesis, viz. that

> "Representation is the wrong unit of abstraction in building the bulkiest parts of intelligent systems"
>
> Brooks, 1991, p. 140

We may already note that Brooks here moves from a lesson concerning the non-necessity of explicit representations (first quote) to a much more general hypothesis concerning the role of representation in general (second quote). By way of a historical aside, we may also note that Heidegger's arguments are likewise best taken as targetting *explicit* representation first and foremost (See Dreyfus, 1991, p. 4). To do justice to Brook's position, however, we must first set it in the context of some practical results.

Brooks set out, true to the guiding philosophy outlined earlier, to build whole simple intelligences capable of performing some task in real-time in a real 'messy' environment. In order to achieve this robust, real-time response, he found it helpful to reject what he terms "the strongest traditional notion of intelligent systems", viz. the notion of

> "a central system, with perceptual modules as inputs and action modules as outputs (in which) perceptual modules deliver a symbolic description of the world and ... action modules take a symbolic description of desired actions and make sure they happen in the world"
>
> Brooks, 1991, p. 146

According to this traditional model, inputs are transformed into a symbolic code which is then the object of computational manipulations in central processing. At the heart of Brooks alternative is an attempt to sidestep the need for a symbolic interface. In place of the familiar functional decomposition (into peripheral input systems, central systems and output systems) Brooks proposes a activity-based decomposition. The simple, whole intelligent system (a 'creature') will comprise a variety of behavior-producing subsystems. These do not, however, act so as to send a symbolic re-coding of

the input to some central system which has then to decide what to do. Rather, each such behavior-producing subsystem (a 'layer') will itself constitute a complete path from input to action. The trick then lies in the orchestration of these relatively independent resources -an orchestration most easily achieved by building in relations of inhibition and override between the various layers. This kind of overall set-up Brooks dubs a 'subsumption architecture' insofar as it consists crucially of self-standing behavior-producing layers (thus allowing incremental testing and design) capable of subsuming each other's operations in virtue of relations of suppression and inhibition whereby the activation (by external input, usually) of a given layer can inhibit the activity of other layers.

For example, Brooks first mobile robot comprised three layers. The first layer was devoted to object avoidance. It exploited a sonar sensing device (in fact, a ring of 12 ultrasonic sonars) whose output was converted to polar co- ordinates and passed to various finite state machines (FSM'S). One such machine was set up to respond if an object is dead ahead by sending a halt signal to a further FSM in charge of the robot's motion. A second FSM computed the presence of other objects and was used to orient the robot to a safe (i.e., unblocked) direction. The upshot of the combined operation of this set of FSM'S was that the robot moved away if you approached it, and generally avoided hitting things.

On top of this Brooks added a second behavior-producing layer whose task was to instigate a wandering behavior if object- avoidance was not in the driving seat. The 'wander machine' generated frequent random course headings which were factored in as a kind of attractive force which was then balanced against the repulsive face 'exerted' by objects to be avoided. The robot thus heads as closely to the random target location as is compatible with object avoidance.

Finally, a third layer (the 'explore machine') can suppress the activity of the wander machine and set up a distant target to be reached, while still exploiting the lower-level strategies so as to avoid local obstacles.

The main point to notice about this architecture (and subsumption architectures in general) is that there is no central system, and no central representations or central representational code. Instead, the creature is just "a collection of competing behaviors" (Brooks, 1991, p. 149). These behaviors are under local environmental control. The system has no central model of its world. In place of such a central guiding model are the more-or-less self-standing layers. These layers "extract only those aspects ... of the world which they find relevant" (Brooks, op. cit., p. 148). Above all, there is no explicit

representation of goals which is consulted by some central process whose task is to decide which goal to pursue next.

In addition to this demonstrable lack of central systems, Brooks further claims that there is no representation involved here, even at the local level of the various behavior-producing layers and finite state machines. This is so, he claims, because:

> "We never use tokens which have any semantics that can be attached to them. The best that can be said in our implementation is that one number is passed from a process to another"
>
> Brooks, 1991, p. 149

In related vein, in contrasting his approach with that of the connectionist, Brooks comments that:

> "Connectionist seem to be looking for explicit distributed representations to spontaneously arise from their networks. We harbour no such hopes because we believe representations are not necessary and appear only in the eye or mind of the observer"
>
> Brooks, 1991, p. 154

We shall return to these rather more problematic claims later on. For now, we may summarize the Moboticist ethos as involving:

1.Scepticism concerning classical task decomposition and the central process/peripheral input process distinction.

2.A commitment to the use of the world, where possible, as its own best 'representation'.

3.The achievement of integrated intelligent behavior via a subsumption architecture, i.e., without any central system managing goals, sub-goals, etc.

4.The rejection of the need for representations of the world, even at the level of individual behavior-producing layers.

3. More Insects, and James Watt.

The moboticist movement is not alone in its opposition to representationalism. More recently, a variety of studies involving genetic algorithms, Artificial Life and Dynamical Systems Theory (more on all of which below) have likewise added their voices to the anti-representationalist clamour. The best way to get the flavour of this increasingly unified approach is to review, in a little detail, one research project which brings them all together: Randall Beer's attempt to evolve walking behaviors in a six-legged 'autonomous agent'. The interest of this project, for our purposes, will lie largely in the systematic attempt to replace the basic explanatory framework of contents and representations with the potentially representation-avoiding framework known as Dynamic Systems Theory. In addition, we shall see some interesting pressure put on the role of computation itself.

In seeking to evolve an autonomous agent, Beer is consciously echoing the basic mobotics ethos outlined above. The goal is to produce an actual robot capable of remaining in robust, long-term interaction with a real environment. In pursuing this goal, Beer too is led to question dominant conceptions concerning the role of representation in intelligent activity. In particular, he questions what he calls the 'computational theory of cognition'. This theory is depicted as claiming that:

> "an agent behaves 'intelligently' in its environment only insofar as it is able to represent and reason about its own goals and the relevant properties of its environment"
>
> Beer, to appear-a, p. 3

Relative to this notion of a computational theory of cognition (one which builds in notions of representations) Beer is willing to claim that a computational theory is not appropriate for characterizing the behavior of autonomous agents in general. He does not, however, deny either (a) that computer modelling is a useful tool for constructing and understanding autonomous, environmentally- situated agency or (b) that there exists some principled relationship between sensory inputs and behaviors. Instead, the central claim is that in very many cases, we do not need a layer of internal representation (as opposed to mere internal *state* -see below) to intervene between the input and the output. Indeed, the whole idea of cognition as what takes place betweem well-defined episodes of receiving input and delivering output is here seen as part of a traditional and non-compulsory framework for thinking about the mind. The contrast between a system which is genuinely (intrinsically) computational and one which is not genuinely computational but nonetheless admits of computational modelling (e.g. fluid

behavior / planetary motions) is , for Beer, just the contrast between systems which really do form and use internal representations and ones which do not.

In place of the (so-called) computational framework of symbols, representations and semantics, Beer exploits a formalism based on Dynamic Systems Theory. This framework, he claims, is much better suited to autonomous agent research and has the virtue of not begging the question in favour of the hypothesis of internal representations. Dynamic systems theory is at root a formalism for describing and understanding the behavior of complex systems (see e.g. Abraham and Shaw, 1992).

The core ideas behind a dynamic systems perspective are:

(1) The idea of a state space.

(2)The idea of a trajectory, or a set of possible trajectories, through that space.

(3)The use of mathematics (either continuous or discrete) to describe the laws which determine the shape of these trajectories.

The dynamic systems perspective thus builds in the idea of the evolution of system states over time as a fundamental feature of the analysis. As a general formalism it is applicable to all existing computational systems (connectionist as well as classicist), but it is also more general, and can be applied to the analysis of non-cognitive and non-computational physical systems as well.

The goal of a Dynamic Systems analysis is to present a picture of an state space whose dimensionality is of arbitrary size (depending on the number of relevant system parameters), and to promote an understanding of system behaviors in terms of location and motion within that abstract geometric space. To help secure such an understanding, a variety of further constructs are regularly invoked. These constructs capture the distinctive properties of certain points or regions in the space as determined by the governing mathematics. The mathematics typically specifies a dynamical law which determines how the values of a set of state variables evolve through time (such a law may consist, for example, in a set of differential equations). Given an initial state, the temporal sequence of states determined by the dynamical law constitutes one trajectory through the space. The set of all the trajectories passing through each point is called the flow, and it is the shape of this 'flow' which is the typical object of study. To help understand the shape of the flow, a number of constructs are used including, for example, that of an Attractor (a point, or region -set of points- in the space such that the laws governing motion through the space guarantee that any trajectory passing close to that point / region will be 'sucked in' to it). Related concepts include 'basin of attraction'

(the area in which an attractor exerts its influence) and 'bifurcations' (cases where a small change in the parameter values can reshape the flow, yielding a new 'phase portrait' i.e. a new depiction of the overall structure of basins and boundaries between basins -separatrices-).

The dynamic systems approach thus aims to provide a set of conceptual and mathematical tools able to promote an essentially geometric understanding of the space of possible system behaviors. New tools are always to be welcomed, and the dynamic systems perspective is, we believe, a powerful and useful one. Accompanying the tools, however, is a clearly articulated scepticism concerning any representational interpretation of the systems thus described. Such systems, Beer concedes, will be rich in 'internal state' (hence, e.g. they need not respond the same way whenever they receive the same input -a lot can depend upon prior state). But even rich internal state, he stresses

> "... does not necessarily correspond to a *representation* of anything in particular, any more than the concentrations of reactants in an industrial fractionation column serve to represent anything about the company outside. Both systems [dynamic systems with internal state and fractionation columns] simply have time-dependent input-output behavior..."
>
> Beer & Gallagher, 1992, p. 115

To illustrate these ideas, Beer describes a series of experiments (see also Beer, 1990, Beer et. al., 1992, Beer & Gallagher, 1992) in which walking behaviors are evolved in artificial insects. The experiments are indeed fascinating, and demonstrate clearly how we can understand e.g. types of walking gait in terms of movement within a state space. For example, one type of insect leg-movement controller is analyzed using the new tools in terms of a systematic flipping between two fixed point attractors. The first comes into play when a foot has just been put down and a 'state phase' begun. The evolution of this state takes the system to a fixed point attractor. As the leg continues to move, however, this attractor disappears to be replaced by a second attractor elsewhere in the state space towards which the system state then evolves. This second attractor corresponds to a 'swing phase'. The switch between these fixed points occurs due to a set of bifurcations which occur as the leg moves through a certain angle. The effect of this is to switch the phase portrait of the controller between the two fixed point attractors. Beer and Gallagher also showed that controllers could be evolved capable of

operating both in the presence of sensory feedback from the environment and (in a degraded fashion) even in its absence. The latter possibility exploits the ability of the evolved network to oscillate so as to generate the required rhythmical control signals. This is presented (Beer & Gallagher, 1992, p. 115) as an example of how non-representational inner state can buy a degree of independence from ambient environmental stimuli.

Beer's results are certainly interesting, and bear out many of the suspicions of the moboticist. He found no clean functional decomposition of leg controller components, and no obvious explicit representations of environmental states (instead, we see a close coupling between environmental states and system responses). In addition, these walking behaviors do not invoke any 'central control'. Finally, the operation of the controllers could be usefully modelled using the tools of dynamic systems theory rather than those of a computational / representational approach. But none of this, on the face of it, amounts to much in the way of evidence for what we shall now dub the General Radical Claim, viz. the claim that internal representation is not essential to genuine cognition. In the next section we shall review some attempts to construct arguments which would (if successful) go some way towards bridging the gap between the moboticist and / or dynamic systems theory evidence and the general radical claim.

4. From Gizmos to Arguments?

How does the anti-representationalist hope to parley the kinds of results and observations rehearsed above into a general case against internal representation? We shall argue that the links between the kinds of result reported in sections 2 and 3 and the more general anti-representationalist conclusions are surprisingly weak. We can discern just three ways in which the bridging manoeuvre is attempted. Two of these are straightforwardly flawed. The third depends on an unwarranted projection from cases involving simple physically present and simply specifiable parameters to more 'representation-hungry' cases requiring sensitivity to distal, non-existent or highly abstract properties. We can take the three manoeuvres in turn.

The simplest, but most obviously flawed, manoeuvre is to identify representation with explicit representation in a 'cut and paste' computational architecture, and then claim that the latter does not provide for fluent real-time coupling with a changing environment. Thus both Beer and Brooks repeatedly stress that explicitly representing e.g. the environmental state may not constitute a fast, flexible means of engaging with that same environment. Such observations, however, do not rule out the kinds of fast,

efficient coupling often achieved by connectionist neural network style solutions (e.g., Churchland's crab -see P. M. Churchland, 1989, ch. 5. For critical comment, see Van Gelder, in press-a); solutions which are nonetheless recognised as falling into a more generally representationalist camp.

But, if explicitness of representation is not the real issue, what is? Beer suggests (and here we confront the second bridging manoeuvre) that the true dispute concerns the *organizational* claims implicit in a computational / representational story. The essence of any non-trivial computationalist story, he suggests, is a claim concerning the way an agent's internal organization mirrors the "functional states and algorithms of a computation" (Beer, to appear-a, p. 9).

The point about non-triviality bears emphasis. Beer concedes that any physical system has a computational description. But this, he rightly observes, cannot support the representationalist's claim that the mind *really is* a computational device. For hurricanes, chemical fractionation towers, etc, etc all have computational descriptions, yet we do not say that the hurricane (or whatever) is really computing. The mere existence of a computer model of some phenomenon does not prove that the phenomenon itself is computational. The difference, as Beer casts it, is that:

> "Computer models of autonomous agents contain symbolic structures that represent theoretical entities to the modeller, while computational theories of autonomous agents claim that agents contain symbolic structures that represent their situation to *themselves*".
>
> Beer, to appear-a, p. 6.

At first blush, this way of putting things is problematic. For the distinction between an inner state's having representational content for the modeller (only) and its having that content for the system itself is obscure. The full-blooded computational / representational approach is not at all committed to the existence of inner homunculi who read and understand the putatively representational inner items (see e.g. Dennett, 1981). But having given up on inner homunculi, there is no-one except the external modeller to whom the inner structures will *appear* representational. The system just uses the structures; they function within it in a purely causal way. To the extent that our (external, theoretic) best understanding of their cognitive role involves assigning representational contents to them, they are (it seems to us) as full-blooded and genuinely representational as any (non- homuncularist) adherent of a representational / computational theory of mind ever supposed.

Nonetheless, the intention behind Beer's formulation is clear enough. He seeks to distinguish e.g. the computational interpretation of calculators (which is also literally true) from that of e.g. hurricanes (which is literally false). And it is in pursuit of this distinction that he suggests that computational / representational stories are literally true just in case the agent's internal organization mirrors the structure of a computational story and it is in virtue of this mirroring that the system "behaves the way that it does" (Op. cit., p. 9). The latter caveat is necessary since we can always generate a computational story which *does* capture the structure of inner events. The question is, does such a story both (a) capture that structure and (b) give us unique explanatory leverage regarding the system's behaviors?

Beer's claim, then, is that the computational approach provides such leverage only when systems

> "have reliably identifiable internal configurations of parts that can be usefully interpreted as representing aspects of the domain in which the system operates and reliably identifiable internal components that can be usefully interpreted as algorithmically transforming these representations so as to produce whatever output the system produces from whatever input it receives"
>
> Beer, to appear-a, p. 10

With this formulation we have no quarrel. But we think that Beer significantly underestimates the extent of the literal applicability of computational stories to events in the brain and nervous system, thus construed. The root of this underestimation, we suggest, is closely akin to the conflation of explicit representation with representation in general. It is the identification of the notion of representationality within the nervous system with the notion of *symbols* which act as the vehicles of representational content.

The notion of a symbol (see Clark, 1992, 1993) brings with it much of the baggage of traditional (symbolic!) A.I. It invites us to think of well-individuated inner items which carry familiar world-referring contents ('dog', 'cat') and which are manipulated (read, copied, concatenated) by an independent processor. When Beer writes that a fundamental aspect of computationalist stories is that they invoke "the idea that symbols somehow encode or *represent* information relevant to behavior" (Op. cit., p. 7), he may, we believe, be falling prey to the temptation to conflate representationalism in general with the more restrictive (and less plausible) image of a *symbol-manipulating* inner economy.

Interestingly, Beer concedes that a computational approach looks fruitful as a means of understanding the mammalian visual system. But he bases this concession on the evidence that that system is "at least partly decomposable into richly interconnected but somewhat distinct functional modules" (Op. cit., p. 11). What Beer fails to appreciate is that it is not modularity *per se* which *legitimizes* the computational story here. Rather, it is just the fact that incoming information is divided into distinct signals (carrying different types of information -e.g., about shape, color and motion. See Livingstone & Hubel, 1987, Corbetta et al., 1991) and that the routing, transforming and efficient integration of those signals succumbs usefully to a computational depiction. Such a depiction is at least modestly *representational* since (1) it involves the semantic interpretation of the kinds of information carried by different signals and (2) the strategies of routing and integration are evolved precisely so as to enable the overall system to track and respond to salient objects and states of affairs in its world (note that this *already* differentiates such systems from the concentrations of chemicals in fractionation columns, etc).

Beer's response to this will probably be that these (modest) notions of computation and representation may be stretching the bounds of our understanding of both. In this vein, he notes that attempts to superimpose the computationalist vision on highly analog and/or distributed processes may well be

> "pushing a language founded on the step-by-step manipulation of discrete symbols by functionally distinct modules past the breaking point"
>
> Beer, to appear-a, p. 11

We disagree. The language of computation and representation, it seems to us, is a much more flexible tool than Beer suggests. It is already clear that many working connectionists see themselves as steeped in both computation and representation, albeit computation and representation of novel and interesting kinds (see e.g., Smolensky, 1988, Elman, 1991). And the neuroscientists use of representation-talk has arguably *always* (or at least usually) carried significantly less overtones of moveable symbols in a read / write architecture than that of the traditional A.I. researcher. These uses are certainly not *obviously* mistaken, and we think Beer needs to do a lot more to convince us that a fruitful rapprochement between a modestly representational computationalism and the extra insights provided by Dynamic Systems Theory is not on the cards.

So far, then, we have introduced a notion of modest representation and argued that cognizing systems with what Beer labels 'rich inner state' may be better treated as *loci* of modest representations (but not fully-fledged symbols). What the committed

anti-representationalist thus really needs to do is to convince us that conformity to the modest notion just rehearsed is really not enough to establish the truth even of the general representationalist vision. One way to do this would be to argue that although a representational gloss is sometimes possible, that gloss fails to illuminate the essential nuances of the real organism/environment coupling. Just such an argument has recently been developed (van Gelder, in press-b), and it is to this third and final manoeuvre that we now turn.

Van Gelder seeks to convince us that the image of 'cognition as computation' is no longer the 'only game in town'. Instead, there is cause (he claims) to take very seriously an alternative notion viz that "cognition is state-space evolution in certain kinds of non-computational dynamical system" (van Gelder, in press-b, p. 1). This turns out to build in some degree of anti-representationalism since it transpires (see below) that computational solutions are distinguished, at least in part, by their reliance on internal representations.

Much of the weight of Van Gelder's exposition is borne by a central example of a non-computational dynamical system, viz, the Centrifugal Governor (see below). For the claim he seeks to highlight is that "cognition itself might be the behavior of dynamical systems relevantly similar to the Centrifugal Governor" (op. cit., p. 36). What, then, IS the Centrifugal Governor?

The governor was designed by James Watt in the late 18th century as a solution to the problem of keeping constant the speed of a flywheel to which machinery is connected. The speed of the flywheel varies according to the steam fluctuations that take place in the engine workload and the boiler. In order to control the speed of the flywheel, we have to control the amount of steam entering the pistons from the boiler via a valve, the so-called throttle valve. What a governor does is to close the throttle valve as the flywheel speed increases -so the flow of steam is restricted- and to open it as the flywheel decreases -letting more steam flow-. In this way the speed of the flywheel is kept constant.

But, what is interesting is not so much what the Watt governor does, but the way in which it does it. When trying to accomplish a difficult task, one strategy is to decompose the task into simpler component tasks which can be performed using available resources. The minimal number of steps into which the steam engine problem can be broken down seems to be the following:

1. Measure the speed of the flywheel.
2. Compare the actual speed against the desired speed.
3. If there is a discrepancy, then:

a. Measure the current steam pressure.

b. Calculate the desired alteration in steam pressure.

c. Calculate the necessary throttle valve adjustment.

4. Make the throttle valve adjustments.

Return to step 1.

(van Gelder, in press-b, p. 2)

What is interesting about Watt's design is that the governor achieves the same aim via a rather different strategy. Watt's solution

> "consisted of a vertical spindle geared into the main flywheel so that it rotated at a speed directly dependent upon that of the flywheel itself. Attached to the spindle by hinges were two arms, and on the end of each arm was a metal ball. As the spindle turned, centrifugal force drove the balls outwards and hence upwards. By a clever arrangement, this arm motion was linked directly to the throttle valve. The result was that as the speed of the main wheel increased, the arms raised, closing the valve and restricting the flow of steam; as the speed decreased, the arms fell, opening the valve and allowing more steam to flow. The result was that the engine adopted a constant speed, maintained with extraordinary swiftness and smoothness in the presence of large fluctuations in pressure and load."
>
> (van Gelder, in press-b, p. 3).

The importance of the way in which the governing problem was solved, van Gelder states, is that the task is performed without any representation of the speed of the flywheel or the throttle valve adjustments. He advances four different arguments in support of the non-representational character of the governor. The first notes that the mere fact that there is some causal correlation between the arm angle and the engine speed cannot all by itself suffice to make the former a representation of the latter. The universe is stuffed with correlations and it is implausible to count them all as representations (think of accidental correlations). We agree, but note that the correlations between eg. specific brain states and color perception look to fall onto the intuitively acceptable side of such a divide. The second argument denies the explanatory utility of treating the governor as a representation-manipulating device.

We agree -there is indeed no extra 'explanatory leverage' gained by so doing. The third argument points out, interestingly, that the governor can even function during transitional moments in which engine speed/arm angle correlations are disrupted. We think this is a powerful property but (as will soon become clear) that other problems with the Governor illustration undermine its impact. The fourth, and final, argument is that the notion of representation itself is simply not rich enough to capture the highly complex dynamical interaction between the arm angle and the engine speed. The behavior of the Watt governor is the result of subtle and constant influences between the arm angle and the engine, i.e., the behavior of the system can only be fully understood as the result of the complex interactions between the physical structure of the system itself and the nature of the environment that surrounds it -i.e., the engine-:

> "A representation is, roughly, an entity that we can see as standing for some other state of affairs. Clearly, however, the arm angle does not stand for the engine speed in any straightforward sense; their relation is much more subtle and complex. The arm angle fixes the way in which the engine speed changes and the engine speed fixes the way in which the arm angle changes -or, more precisely, it fixes the way in which change in the arm angle changes, depending on what that arm angle happens to be"
>
> (van Gelder, in press-b, p. 18).

The claim, then, is that the coupling between the governor and its 'environment' (the steam engine) is so tight and complex that treating one system as representing states of the other gives us no adequate explanatory purchase on the system's behaviors. Instead, we will understand the space of behaviors best by focusing on the dynamical law which it instantiates. van Gelder thus goes on to rehearse this law and then to tell a story in the now-familiar vocabulary of dynamic systems talk concerning how e.g. given engine speeds determine locations for a fixed point attractor in the overall state space. Increases in engine speed shift the location of the attractor and it is this shifting which ultimately ensures that the desired flywheel speed is maintained. The full story is complex and indeed satisfying. But the general conclusion (that representational analysis are too superficial to figure in the full understanding of the behavior of complex coupled dynamical systems like agents and environments) seems radically underargued. Here's why.

The basic trouble is one that afflicts all the case studies mentioned above. It is that the kinds of problem-domain invoked are just not sufficiently 'representation-hungry'. Instead they are, without exception, domains in which suitable ambient environmental stimuli exist and can be pressed into service in place of internal

representations. We agree that it is important that the Cognitive Scientist be aware of the potential for 'shortcuts' which nature provides (see e.g., Clark, 1989, ch. 4). But it is unfair to use these cases to illustrate any more general anti-representationalist claim.

By a 'representation-hungry' problem domain we mean any domain in which one or both of the following conditions apply:

1.The problem involves reasoning about absent, non-existent, or counterfactual states of affairs.

2.The problem requires the agent to be selectively sensitive to parameters whose ambient physical manifestations are complex and unruly (for example, open-endedly disjunctive).

The first class of cases will be familiar enough. The ability to track the distal or the non-existent requires, prima facie, the use of some inner resource which enables appropriate behavioral co-ordination without constant ambient input to guide us. Whatever plays that kind of inner role is surely going to count (see our comments on Beer above) as some kind of internal representation. We note only that the ability to reason about the absent, non-existent or counterfactual is not plausibly dismissed as merely a 'tip of the iceberg' cognitive phenomenon. Non-language using animals (e.g., chimps hunting in packs) seem to anticipate the movements of pursued prey and to engage in counterfactual reasoning. A nice and well-documented example of the latter concerns grooming behaviors in rhesus macaques (monkeys). These animals seem able to make quite sophisticated judgements concerning the motivational states of their peers. In combat situations, support from a high ranking female is often decisive. Monkeys who groom such females tend to receive such support. Hence, it is wise to avoid contests with macaques who have been seen grooming these females. Such avoidance behavior is indeed often found, and persists long after the visual stimulus (witnessing the grooming event) has ceased. Knowledge of the likely behavior of the high-ranking female in combat situations that have not yet arisen thus seems essential to the social organization of the group (see Harcourt, 1985). Yet a good explanation of such behaviors will *prima facie* need to acknowledge some kind of internal representation of positions in the social hierarchy, and storage in memory of knowledge concerning past grooming events.

The second class of cases is equally compelling, though a little hard to describe. The idea is that much cognition involves the development of sensitivities to states of affairs whose manifestation in the sensory inputs is, to say the least, attenuated. These are states of affairs which are highly relational or otherwise functional and 'abstract'. The ability to respond selectively to all and only the *valuable* items in an array, or to all

and only those items which belong to the Pope would be cases in point. It is hard, on the face of it, to see how to set up a system to track such properties unless it is capable of subsuming a variety of superficially very different inputs under some common rubric, and *then* defining further processing events not over the sensory array but over some inner item or pattern whose content corresponds to the more abstract property in question. Yet to do this is, surely, just to invoke the idea of an internal representation.

It will be objected that sensitivity to such features as valuableness surely *is* a 'tip of the iceberg', language-user specific, phenomenon. But the underlying point is in fact much more general. It is that behavioral success often depends on the ability to *compress or dilate* an input space. That is, the cognizer must be able to treat inputs whose immediate codings (at the sensory peripheries) are quite similar as deeply different (dilation) and conversely, to treat inputs whose immediate codings are quite different as deeply similar (compression). On the modest reading which we recommend, internal states developed to serve this end just *are* internal representations whose contents concern the states of affairs thus isolated. Such internal states function as feature detectors and enable the system to differentially respond to situations for which it has no dedicated transducer-level detection mechanisms. It can be shown that the successful negotiation of even fairly simple looking toy domains often depends crucially on the use of such strategies -see Clark & Thornton, submitted-.

Recent neuroscientific research has also argued that basic visual abilities (such as object recognition) may require the use of related strategies. The ability to recognize the same object from any one of a number of distances, angles, settings, etc is best explained, this research argues, by supposing that the system first transforms the input into a canonical presentation frame (with position and scale invariant) and only then matches this transformed product to its stored knowledge so as to carry out the identification task. Such a strategy is steeped in computation (for the input transformation) and representation (for the matching), and yet is invoked to explain basic visual abilities common to many animals (for full details, see Van Essen et. al., in press).

The same authors end by developing an image germane to our concerns. Models of cortical function, they claim, should treat the brain as

> "... a system designed to treat information as an essential commodity, much as an efficient factory is designed for optimal handling of the physical materials that traffic across its floors. In both cases the raw materials that enter the system generally represent only a small fraction of the final product. The production process involves careful selection of

> useful materials, discarding of excess or unnecessary materials, and transforming and repackaging of the desired materials in an appropriate configuration for the particular applications for which the product is intended."
>
> Van Essen et. al., in press, p. 28

The products of such processes of selection, transformation and repackaging are, we claim, modest internal representations: internal information bearing states which capture regularities which are not available in the simple surface statistics of the input arrays, but instead emerge only as a result of the subsequent filtering and transformation of such signals.

The point we wish to make, however, is *not* that it is simply inconceivable that object recognition, counterfactual reasoning and selective response to rather abstract kinds of features might all succumb to some unexpected, representation-free, kind of explanation. Rather, it is that *insofar* as the robot insect/governor style cases are meant to illustrate the tenability of a general anti-representationalism, they seem to us to miss the mark. For the problem domains they negotiate are not (yet, at least) the ones on which the representationalist should rest her case.

Consider, in this context, Skarda and Freeman's impressive (1987) model of the olefactory bulb. Skarda and Freeman here provide a beautiful and challanging Dynamic Systems model of the way sensory information is registered in the olefactory bulb. But they later claim that this model lends support to a much grander claim viz that

> "The concept of 'representation' ... is unnecessary as a keystone for explaining the brain and behavior [because] the dynamics of basins and attractors can suffice to account for behavior without recourse to mechanisms for symbol storage".
>
> Skarda and Freeman, 1987, p. 184

But the evidence they present, it seems to us, comes nowhere close to providing support for such a weighty claim, since it is not addressing a truly 'representation-hungry' problem type. (Note once again the persistent conflation of the general notion of representation with more specific ideas such as 'symbol storage').

What the various cases discussed above show is thus at most that:

> (1)The conceptual apparatus of Dynamic Systems Theory provides a powerful tool for understanding the behavior of complex, coupled dynamical systems.

And

(2)that insofar as the dimensions which define the relevant state spaces are *themselves* best understood as reflecting simple physical properties detectable in the ambient input, *then* the Dynamic Systems story we need will be a non-representational one.

In the case of the Governor, it is no surprise at all that the best story is told in non-representational terms. Likewise for a multitude of other systems whose tasks require responses only to physical parameters available without undue computational effort in the ambient environmental stimuli. The case of the insect leg controller is only marginally more surprising. True, someone just might have thought that the best solution here involved representing the environment and / or goals. But it also seems quite natural to skip the representational intermediaries and model the leg controller as merely a well-coupled dynamical system. Likewise for most (all?) other examples in this literature (e.g. the model of tone-intensity storage describe in van Gelder, op. cit., p. 20, and all the moboticist work introduced in Section 2).

It is worth reminding ourselves, at this point, that the tools of dynamic systems theory are not in any way intrinsically non- representational. The dynamic system's style of analysis and understanding can just as well be used as part of a thoroughly representation-laden story. The determining factor is just the nature of the *dimensions* of the relevant state space(s). The more distance there is between the state of affairs to be tracked and the bare physical parameters detected by the transducers (the more success depends on dilating and compressing the input space, on filtering and tranforming superficially different inputs to reveal deeper commonalities), the more we will be inclined to view motion within the state space as (precisely) motion within a high-dimensional representational space. By telling Dynamic Systems Stories about state spaces whose dimensions are specified by straightforwardly physical properties, one can make it look as if all rich dynamic system's explanations will constitute replacements for representational ones. But this is far from the case, as evidenced by the growing body of connectionists who use dynamic systems constructs as part of a thoroughly representational approach. Thus Paul Churchland depicts connectionist networks as essentially embodying knowledge structures organized around *prototype-style* representations. He then goes on to depict the prototype as "a point or small volume in an abstract state space of possible activation vectors" (P. M. Churchland, 1989, p. 206) and highlights the geometric relations obtaining between the various prototype representations constituted within a single space. In describing the functionality of the prototype representation he states that:

> "In dynamical terms, the prototype position is called an 'attractor'. We may think here of a wide mouthed funnel that will draw a broad but delicately related range of cases into a single narrow path".
>
> P. M. Churchland, 1989, p. 206.

To take just one more case, consider the following extract from a recent study in which lesioned neural networks were used to simulate the reading errors produced after certain kinds of brain damage:

> "We have found it useful to think of the networks output ... as motion through a multidimensional semantic space ... [the] notion of a semantic space dotted with attractors representing the meanings of words has proved valuable for understanding how our network operates ..."
>
> Hinton, Plaut & Shallice, 1993, pp 79/80.

The success of a non-representational analysis of a device like the Watt Governor thus fails to argue for any more generic anti-representationalism. For since the dimensions of the relevant state space were straightforwardly physical (available without significant computational effort from the ambient environmental input), the result is effectively trivial. By contrast, as soon as we are dealing with state spaces whose dimensions are more abstract, and hence cover a superficially very disparite range of patterns of physical stimulation (as in e.g. responding to an item as 'valuable', or even detecting the presence of a given phoneme (see Seidenberg & McClelland, 1989), the dynamical system story becomes a representational one (too). Thus, unless you believe that human cognition somehow operates without re-coding gross sensory inputs so as to draw out the more abstract features to which we selectively respond, you will already be committed to a story in which the state spaces themselves are properly seen in representational terms.

It is worth recalling at this point Brooks' description (section 2 above) of behavior producing layers as extracting just "those aspects ... of the world which they find relevant" (Brooks, 1991, p. 148). For it is our contention that *as soon as* we are forced to acknowledge any *internally* driven sorting, filtering or transforming of inputs then we are immediately dealing in modestly representational state spaces. Where very low level behaviors are concerned, it may indeed be viable, as Brooks suggests, to have the filtering effectively done *by the transducers* (e.g., in a given mode, the sensors will only respond to certain types of input). But as the level of abstraction of the properties to be detected increases, and as the multiplicity of different ways the incoming data must be used increases, it becomes pretty much mandatory to engage in a variety of processes

in which the data is filtered, transformed, separated into channels carrying different kinds of information etc. (It is, for example, an unusually clear lesson of neuroscientific research (see e.g., Churchland & Sejnowski, 1992, Van Essen et. al., in press) that sensory inputs to e.g. the eye are *subsequently* channelled and transformed so as to allow distinct neural systems to compute e.g. shape and location. In seeking to understand the flow and use of information in the brain it thus appears to be mandatory to identify a variety of state spaces whose dimensions are to be understood not merely numerically but in terms of the different kinds of information which they encode).

Our view, then, is that the radical dynamic systems claim (that cognition is "state-space evolution in Watt-type systems" (van Gelder, op. cit., p. 37)) is unsupported. What work like that of van Gelder, Brooks and Beer really shows is just that the smaller the gap between the dimensions of the relevant state space and bare, easily available input parameters, the less the need to indulge in the more *representationally-infected* types of dynamic systems talk. As the gap increases (as information is filtered and re-coded through a cascade of layers of processing), so too the appropriateness of a representational depiction of the dimensionality of the state space increases. What is genuinely novel, it seems to us, is the dynamic systems style depiction of key processing events in terms of trajectories through a state space. This is a rich notion, and one which opens up ways of thinking about how representations function which are quite unlike the classical vision of symbol copying and combination (see also Clark, 1993). It is our understanding of how representations are processed and transformed, and not our understanding of representation per se which should be the true epicentre of the dynamic systems challenge.

5. Revisionary Representationalism.

It will be useful to end by re-constructing the dialectic as we see it, and making some more positive comments. We begin by noting that Van Gelder, like ourselves, holds out hope for a kind of revisionary representationalism. Thus he notes that some connectionist work (e.g Smolensky, 1988, Port, 1991, Elman, 1991) fits rather nicely into the general Dynamic Systems framework. Van Gelder even goes so far as to explicitely counter the idea that he aims to oppose Representationalism *tout court*. Thus he comments that:

> "The centrifugal governor is not representational, but I am not suggesting that representation has no role. In fact, an exciting feature of the

dynamical perspective is that it opens up dramatic new ways of thinking about representation".

Van Gelder, in press-b, p. 36

He also concedes that everything he suggests is entirely compatible with "*some* aspect of a system being representational" (op cit, p. 37, original emphasis). What are we to make of these concessions?. On the one hand, they seem to place him not (after all) in the anti-representationalist camp but (with us) in the revisionary representationalist one (see also Van Gelder & Port, 1993). On the other hand, to position Van Gelder thus is to render obscure the role of his extended treatment of the Centrifugal Governor. For it is central to that treatment that he sets out to convince us:

1.That the Governor works without exploiting representations of any kind, and

2.That cognition itself may be best understood in terms of the operation of dynamical systems "*relevantly similar* to the Centrifugal Governor" (op. cit., p. 12, our emphasis), and

3.that cognitive systems are dynamical systems "found *relatively close* in the space of possible systems to the Watt Governor" (op. cit., p. 36, our emphasis).

It is unclear how to interpret these claims of conceptual proximity to the Watt Governor except by supposing that the proximal systems share the characteristic which Van Gelder was at such pains to establish (using no less than 4 distinct arguments) viz we must surely assume that the proximal systems likewise succeed by exploiting close couplings achieved without the benefit of representational intermediaries. Or must we? The aim of the example, we are later told, is just to gain maximum contrast with a Cartesian (representationalist) world view so as to force a radical reconsideration of all its elements *before* allowing the re-incorporation of some of them into a dynamical perspective (op. cit., p. 37). But to the extent that any such re-incorporation vindicates a notion of internal representation, it is surely incompatible with the strong claim that cognitive systems lie close in conceptual space to the paradigmatically non-representational Watt Governor. Surely something has to give. Either the concessions to representationalism are not what they seem or the claim that cognition is state space evolution in systems conceptually close to the Watt Governor is not supposed to be believed.

One way to repair this apparent tension is to focus on a quantitative question : How *much* of what we think of as cognition requires the exploitation of internal representations? One way of reconciling the surface tensions just described is thus to

cast the claim concerning proximity to the Watt Governor as asserting that an unexpectedly large portion of what we generally take as our cognitive achievements may in fact turn out to be explicable without the invocation of any kind of internal representation. The residue, however, may require explanation in terms which re-incorporate at least some aspects of the classical notion of internal representation. Such a position would be vindicated if, for example, it turned out that all aspects of our cognitve achievement *except* the capacity to reason about the non-existent and the spatio-temporally distant could be explained in Governor-style terms. A proper response, were things to turn out thus, would be that insofar as the bulk of our cognitive activity is rather defined in terms of 'perceiving' and 'acting', and these (it seems) have been conceded as susceptible to non-representation invoking explanation, the moral victory goes to the Anti-Representationalist. And, as Tim van Gelder (personal communication) has pointed out, this would be a significant victory indeed, since it would unseat the dominant conception of perception as a process whereby internal representations are formed, and action as the carrying out of internally formulated commands.

Our claim, however, has been that the 'exotic' achievements involving reasoning about unicorns, prime numbers, etc. by no means exhust the class of 'representation-hungry' problems. In fact, that class turns out to embrace all the more mundane episodes in which we are able to engage in fast, on-line counterfactual reasoning, or to respond to perceived inputs in ways which depend on identifying them as instances of fairly abstractly defined states of affairs. For to do so (for us to identify a Renoir as deserving careful treatment in virtue of its *valuableness*, or, for a monkey, to identify an otherwise unremarkable peer as deserving careful treatment in virtue of the rank of the females it has recently groomed) requires (*prima facie*) the ability to subsume an open-ended disjunction of physically specified inputs under some common rubric which enters into the determination of our responses. Any process in which a physically defined input space is thus transformed so as to supress some commonalities and highlight others is, we claim, an instance of modest representation. The greater the computational efforts involved in effecting the tranformations (generally, the more distant the target features are from the first order statistics of the input. See Clark and Thornton, submitted), the *more representational* the solution.

On our account, the notion of Representation is thus re-constructed not as a dichotomy but as a continuum. At the non-representational end of that continuum we find cases in which the required responses can be powered by a direct coupling of the system to some straightforwardly physically specifiable parameters available by

sampling the ambient environment in some computationally inexpensive way (eg. a toy car with a 'bump' sensor). Moving along the continuum we start to find cases in which the system is forced to dilate and compress the input space : to treat as similar cases which (*qua* bare input patterns) are quite unalike and to treat as different cases which (*qua* bare input patterns) are pretty similar. At this point, the systems are trafficking in 'modest representations'. In addition, where such dilation and compression is achieved by the creation of a systematically related body of intermediate representations (as in the connectionist learning of such representations at the hidden unit level (see eg., P.M.Churchland, 1989) we begin to witness the emergence of full-blooded representational 'systems', albeit ones which remain quite unlike the classical vision of such systems as loci of *moveable* symbols capable of literal combination into complex wholes. (It is this classical vision of moveable symbols prone to engage in text-like recombinative antics which corresponds most closely to the vision of *explicit* representation which is, we claim, the proper target of many of the 'anti-representationalist' arguments). And at the far end of the continuum we find cases in which the system is able to invoke various kinds of intermediate representations even in the absence of ambient environmental stimuli (ie as a result of 'top-down' influences.). At this point, we find systems capable of reasoning about the spatio-temporally remote etc.

If the continuum picture is on the right track, then a respectable notion of internal representation will very probably be forced upon us long before we reach the level of cognitive sophistication marked by the truly 'exotic' and linguistically infected cases. It is this claim which most clearly distinguishes our position from Van Gelder's own (and, we think, from that of the other 'anti-representationalists' treated above). At the very least, it is our claim that nothing in the stories about robots, insects, Watt Governors etc. puts any kind of pressure on the representationalist case thus developed. Such pressure, if it exists at all, is coming not from the practical case studies but from the Ghostly Undertow of classical philosophical opposition to the Cartesian world-view. Heidegger may yet win the day. But mobots won't help secure the victory.

6. Conclusions: The Capacious Toolkit.

Recent developments in real-world robotics and the use of Dynamic Systems Theory to explain the behavior of complex and/or highly coupled systems have been seen as contributing support to a general scepticism concerning the role of internal representation in explaining the bulk of human cognition. On closer examination,

however, these recent developments fail to deliver the anti-representationalist goods. The main cause of this failure is their inattention to what we have termed the 'representation-hungry' kinds of problem domain. These domains extend far beyond 'tip of the cognitive iceberg' tasks (like reasoning about the non-existent) and include a variety of cases of perceptual recognition and perceptually guided action.

It may be that part of the 'anti-representationalists' goal is to cast doubt on the role of *explicit* representations, conceived as moveable tokens capable of entering into literal text-like recombination to express complex contents. Such scepticism is well grounded (see Clark, 1993). But we should not confuse this issue with the more general issue concerning representation itself.

It may also be that the goal is (at times) not so much to doubt the existence and role of internal states properly glossed in representational terms but simply to claim that to conceive of such states PURELY in representational terms is to miss much of their power and interest. Here too, we can safely agree. Much will no doubt be learnt from treating the couplings *between* innner subsystems (and, of course, between the agent and the world) in Dynamic Systems terms. But if (as we have argued) the dimensions of the state spaces at issue are often themselves best conceived in representational terms, then what is at issue is a welcome enrichment of the representationalist story, and not its downfall.

In sum, these are exciting times and there is a lot to learn. Dynamic Systems Theory and real world robotics are among the most promising new tools for Cognitive Science. But for the present, at least, a space for representation in that capacious toolkit remains assured.

ACKNOWLEDGEMENTS

Thanks to Tim van Gelder, Dave Chalmers, William Bechtel, Keith Butler, Morten Christiansen, and Matthew Elton for comments and criticisms of earlier versions of this paper. This paper was partly written while one of the authors (Clark) held a Senior Research Leave Fellowship granted by the SERC/MRC Joint Council Initiative on Human Computer Interaction.

Andy Clark / Josefa Toribio
Philosophy / Neuroscience / Psychology Program
Washington University in St. Louis
Philosophy Department
One Brookings Drive
Campus Box 1073
St. Louis, Mo, 63130, USA
e-mail (A. Clark) andy@twinearth.wustl.edu
e-mail (J. Toribio) pepa@twinearth.wustl.edu

References

Abraham, R. H. & Shaw C. D.: 1992, *Dynamics. The Geometry of Behavior*, Addison-Wesley, Redwood City, Ca, 2nd ed.

Beer, R.: 1990, *Intelligence and Adaptive Behavior*, Academic Press, San Diego.

Beer, R.: to appear-a, 'Computational and Dynamical Languages for Autonomous Agents' in T. van Gelder and R. Port (eds.), *Mind as Motion*, MIT Press, Camb. Ma.

Beer, R.: to appear-b, 'A Dynamical Systems Perspective on Environment Agent Interactions', *Artificial Intelligence*.

Beer, R. & Gallagher, J. C.: 1992, 'Evolving Dynamical Neural Networkds for Adaptive Behavior', *Adaptive Behavior*, **1**, 91-122.

Beer, R., Chiel, H. J., Quinn, R. D., Espenschied, K. S. & Larsson, P.: 1992, 'A Distributed Neural Network Architecture for Hexapod Robot Locomotion', *Neural Comp.* **4** (3), 356-365.

Brooks, R.: 1991, 'Intelligence Without Representation', *Artificial Intelligence*, **47**, 139-159.

Churchland, P. M.: 1989, *A Neurocomputational Perspective*, MIT Press, Camb., Ma.

Churchland, P. S. & Sejnowski, T. J.: 1992, *The Computational Brain*, MIT Press, Camb. Ma.

Clark, A.: 1989, *Microcognition*, MIT Press, Camb., Ma.

Clark, A.: 1992, 'The Presence of a Symbol', *Connection Science*, **4**, 193-206.

Clark, A.: 1993, *Associative Engines*, MIT Press, Camb., Ma.

Cleeremans, A.: 1993, *Mechanisms of Implicit Learning: Connectionist Models of Sequence Processing*, MIT Press/Bradford Books, Camb. Ma.

Corbetta, M., Miezin, F. M., Dobmeyer, S., Gordon, L. S. and Peterson, S. E.: 1991, 'Selective and Divided Attention During Visual Discriminations of Shape, Color and Speed: Functional Anatomy by Positron Emission Tomography', *The Journal of Neuroscience*, **11** (8), 2383-2402

Dennett, D.: 1981, *Brainstorms*, MIT Press, Camb., Ma.

Dreyfus, H. L.: 1991, *Being-in-the-World. A Commentary on Heidegger's Being and Time. Division I*, MIT Press, Camb., Ma.

Elman, J. L.: 1991, 'Distributed Representations, Simple Recurrent Networks and Grammatical Structure', *Machine Learning*, **7**, 195-225.

Fodor, J.: 1975, *The Language of Thought*, Crowell, New York.

Fodor, J.: 1987, *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*, MIT Press, Ca. Ma.

Fodor, J. & Pylyshyn, Z.: 1988, 'Connectionism and Cognitive Architecture: A Critical Analysis', *Cognition*, **28**, 3-71.

Giunti, M.: 1992, *Computers, Dynamical Systems, Phenomena and the Mind*, PHD Thesis, Indiana University.

Harcourt, A.: 1988, 'Alls Fair in Play and Politics', *New Scientist*, **108**, 35-42.

Haugeland, J.: 1991, 'Representational Genera', in W. Ramsey, S. Stich and D. Rumelhart (eds.), *Philosophy and Connectionist Theory*, Erlbaum, New Jersey, pp. 61-90.

Heidegger, M.: 1962, *Being and Time*, Harper & Row, New York.

Hinton, G., Plut, D. & Shallice, T.: 1993, 'Simulating Brain Damage', *Scientific American*, **269**, (4), 76-82.

Livingstone, M. S. & Hubel, D. H.: 1987, 'Psychophysical Evidence for Separate Channels for the Perception of Form, Color, Movement, and Depth', *The Journal of Neuroscience*, **7** (11), 3416-3468.

McClelland, J., Rumelhart, D. and the PDP Research Group: 1986, *Parallel Distributed Processing: Explorations in the Micro-Structure of Cognition*, Vol. I and II, MIT Press, Camb., Ma.

Newell, A. & Simon, H.: 1972, *Human Problem Solving*, Prentice-Hall.

Pinker, S. & Prince, A.: 1988, 'On Language and Connectionism. Analysis of a Parallel Distributed Processing', *Cognition*, **28**, 73-193

Port, R.: 1990, 'Representation & Recognition of Temporal Patterns', *Connection Science*, **2**, 151-176.

Rumelhart, D. & McClelland, J.: 1986, 'On Learning the Past Tenses of English Verbs'. in J. McClelland et. al. (eds.), *Parallel Distributed Processing: Exploration in the Microstructure of Cognition*, vol. 2, MIT Press, Camb. Ma., pp. 216-271.

Seidenberg, M. & McClelland, J.: 1989, 'A Distributed, Developmental Model of Word Recognition and Naming', *Psychological Review*, **96**, 523-568.

Smolensky, P.: 1988, 'On the Proper Treatment of Connectionism', *Behavioral and Brain Sciences*, **11**, 1-74.

van Essen, D., Anderson, C. and Olshausen, B.: in press, 'Dynamic Routing Strategies in Sensory, Motor and Cognitive Processing', in C. Koch and J. Davis (eds.), *Large Scale Neuronal Theories of the Brain*, MIT Press, Camb., Ma.

van Gelder, T. J.: 1990, 'Compositionality: A Connectionist Variation on a Classical Theme', *Cognitive Science*, **14**, 335-384.

van Gelder, T. J.: 1991, 'What is the "D" is "PDP"?. A Survey of the Concept of Distribution', in R. W. Ramsey et. al. (eds), *Philosophy and Connectionist Theory*, Erlbaum, New Jersey.

van Gelder, T. J.: in press-a, 'Is Cognition Categorization?', in G. V. Nakamura, R. M. Taraban and D. L. Medin (eds.), *Categorization by Humans and Machines*, Academic Press, San Diego.

van Gelder, T. J.: in press-b, 'What Might Cognition Be if Not Computation?', in R. Port and T. J. v. Gelder (eds.), *Mind as Motion: Dynamics, Behavior and Cognition*, MIT Press, Camb., Ma.

van Gelder, T. J. & Port, R.: 1993, 'Beyond Symbolic: Prolegomena to a *Kama-Sutra* of Compositionality', in V. Honavar and L. Uhr, *Symbol Processing and Connectionist Models in Artificial Intelligence and Cognition: Steps Toward Integration*, Academic Press, San Diego.